An Ontology of Chinese Radicals:
Concept Derivation and Knowledge Representation based on the Semantic Symbols of Four Hoofed-Mammals

Chu-Ren Huang, Ya-Jun Yang and Sheng-Yi Chen
Institute of Linguistics, Academia Sinica
Taipei, Taiwan
{churen, Yajun, eagles}@gate.sinica.edu.tw

Abstract. Semantic symbols are essential components of Chinese characters. ShuoWenJieZi (Xyu Shen 121), the oldest dictionary of Chinese, is organized according to the radical forms as semantic symbols. Characters are classified according radicals, and their meanings cluster around the basic concept of the semantic symbol. We believe that ShuoWenJieZi radicals systemreflect conventional conceptualization when Chinese character orthography was invented. In this research, we use the semantic symbols representing four hoofed-mammals in ShuoWenJieZi ,"bovid," "deer," "cattle," and "horse," as our research objects. In principle we assume that semantic symbols represent basic concepts, and further more we distinguish the relations between derived characters and each basic concept to construct a conventionalized ontology headed by basic concepts expressed by the semantic symbols. Our analysis and comparative studies of the semantic symbol ontologies for the four hoofed-mammals show that they share similar conceptual structures strongly motivated by their functions in human society. In particular, we show that the conceptual dependencies between the basic concept of a radical and the meanings of the derived characters can be explained by an enriched version of the Generative Lexicon.

Keywords: Ontology; HanZi ; Semantic Lexicon

1. Introduction

Chinese radical (yi4fu2,ideographs ;semantic symbols) system offers a unique opportunity for systematic and comprehensive comparison between formal and linguistic ontologies. Previous studies adopt either WordNet-based representation (Wong et al, 2002, and Hsieh, 2006) or SUMO-based mapping (Chou, 2005). Among these studies, Chou and Huang (2007) suggest that the family of Chinese characters sharing the same radical can be linked to a basic concept by Qualia relations. This approach has great implications of accounts for radicals as linguistically conventionalized ontology. In this paper, we take this approach further and try to account for each radical group as domain ontology headed by one basic concept. In particular, we examine in details 4 radicals of animals: 羊 (yang2, bovid) , 鹿(lu4, deer), 牛(niu2, cattle) and 馬 (ma3, horse). Among these four animals, "羊," "牛,"and "馬"are domesticated and serve specific functions in human society. They are highly related to daily lives of human. One of the interesting research issues is to see if the derived concepts of these four animal radicals reflect the differences which show the interaction between animals and human.

Our theoretical foundation is Pustejovsky's Quilia Structure (Pustejovsky, 1995), and the original analysis of 'ShuoWenJieZi'(Xu, 121). In ShuoWenJieZi, all Chinese characters are classified as derived from 540 radicals. In this study, we assume that these radicals each represent a basic concept and that all derivative characters are conceptually dependent on that basic concept. Our study aims at accounting for the exact nature of these conceptual dependencies. Combined with previous work, we suggest that conceptual extensions from the

basic concept encoded by a radical can be classified into seven main types: formal, constitutive, telic, participant, participating, descriptive (state/manner) and agentive.

2. The Semantic Symbol Ontology

The Semantic Symbol Ontology is a system expressing the relations of Hanzi and its meaning cluster.This ontology system extended the basic structure constructed, (Chou, 2005), which maps the meanings of 540 radicals in *ShuoWenJieZi* with IEEE SUMO. We use the results from analyzing derivative concepts to express the Semantic Ontology for each radical. Our current working interface allows easy query of existing database as well as recording of new entries.

2.1.Radical Search

There are two searching methods for the semantic symbol ontology:
(i) Search on SUMO concepts classification
Choose certain SUMO concept, then this concept and its lower SUMO concept will show up on the interface.
(ii) Search on the radical word forms
Key in the radical word form, and users can get the data of that radical directly.

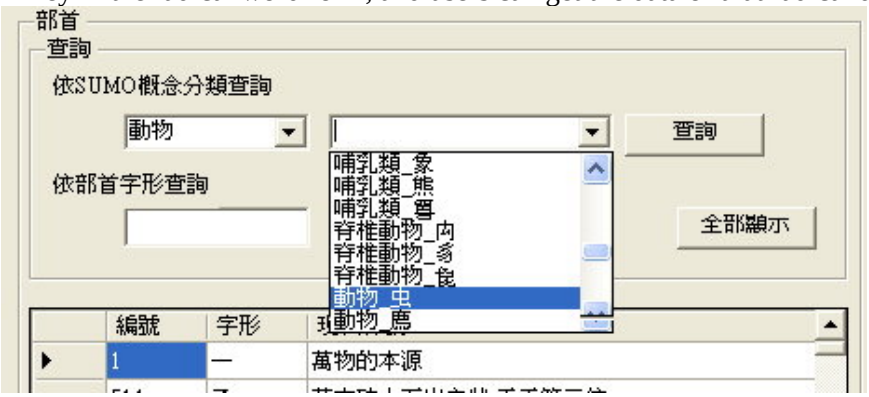

Figure1: Radical search in semantic ontology system

2.2. Basic Concept

According to the definitions in *ShuoWenJieZi* and our analysis of meaning cluster of the characters derived from the same radical, we can posit the basic concept for each radical. For example, the basic concept for "羊" is "mammal with hoof.**"**

2.3. The classification of Hanzi semantic symbols

Based on the definition in *ShuoWenJieZi,* our structure classifies the relationship between deriving meaning cluster and the basic concept of a radical. We use Pustejovsky's *Quilia Structure* as base and observe the analysis on the definitions in *ShuoWenJieZi,* and then classify the deriving concepts of Hanzi radicals into 7 categories, expanded from the orginal four qualia aspects of Formal, Constitutive, Agentive, and Telic:

(i) Formal: This category can be further divided into 5 small categories: "sense," "characteristic," "proper names," and "atypical." The "sense" categories can be further divided into 5 small categories: "vision," "hearing," "smelling," and "taste."
(ii) Constitutive: This category can be further divided into 3 small categories: "part," "member," and "group."
(iii) Telic: Concepts related to function or usage.

(iv) Participant: Words are classified into this category when the definition in *ShuoWenJieZi* mentions the participant involved.
(v) Participating: According to different events, concepts are divided into 6 small categories: “action,” “state,” “purpose,” “function,” “tool,” and “others.”
(vi) Descriptive: This category can be further divided into two categories: “active” and “state.”
(vii) Agentive: The relationship between the radical and its meaning cluster coming from production or giving birth are classified in to agentive.

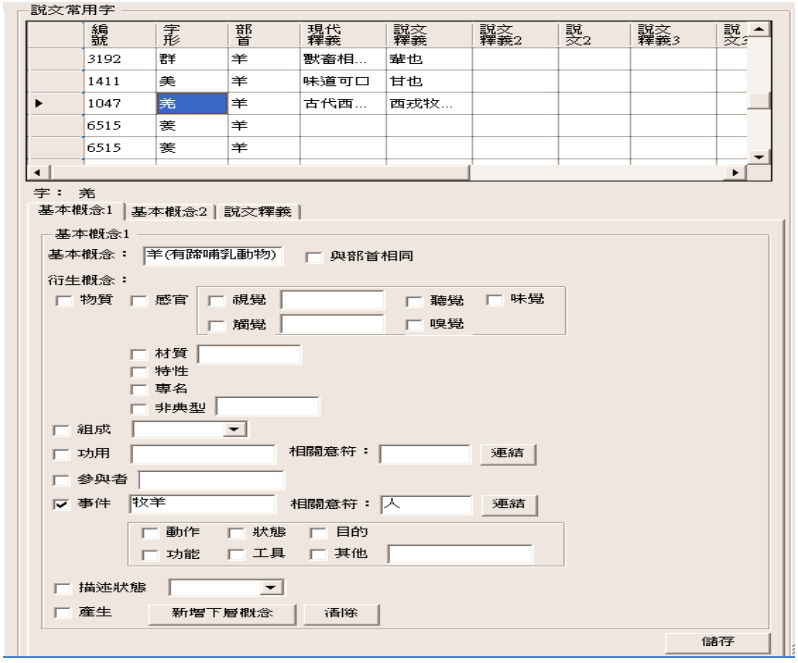

Figure2: The classification of Hanzi semantic symbols

2.4. The link of related semantic symbol

Under “telic” and “participating,” we add a column for “related semantic symbol” to show and link the related deriving concepts. For example, **the character “羌” is explained as “西戎羊橐人也.”(the person whoe herds sheep in Xiyu)** and involves two basic concepts that are represented in the character: bovid and human sincethe Chinese character form of “羌,” contains bothe “羊”（bovid） and“人” (human). Our ontology system links “羌”with its related semantic symbol “人” to offer cross-referencing in oreder to build a more realistic ontology of the conceptual convention.

3. Domian Ontologies of Four Hoofed-Animals as Conventionalized by Radical

3.1. Bovid Domain Ontology Conventionalized by Radical “羊”

According to our analysis, the deriving concepts of “羊” on the category system includes “formal,” “participating,” “constitutive,” “agentive,” and “telic.” Among these five classes, the most prevalent conceptual derivation can be classified as “formal”, which is further classified as Sense, Characteristic and Proper Name by us. Two-thirds of the derived characters denote this conceptual groop..

The following is the concept deriving illustration of radical “羊”, with the top concept of BOVID.

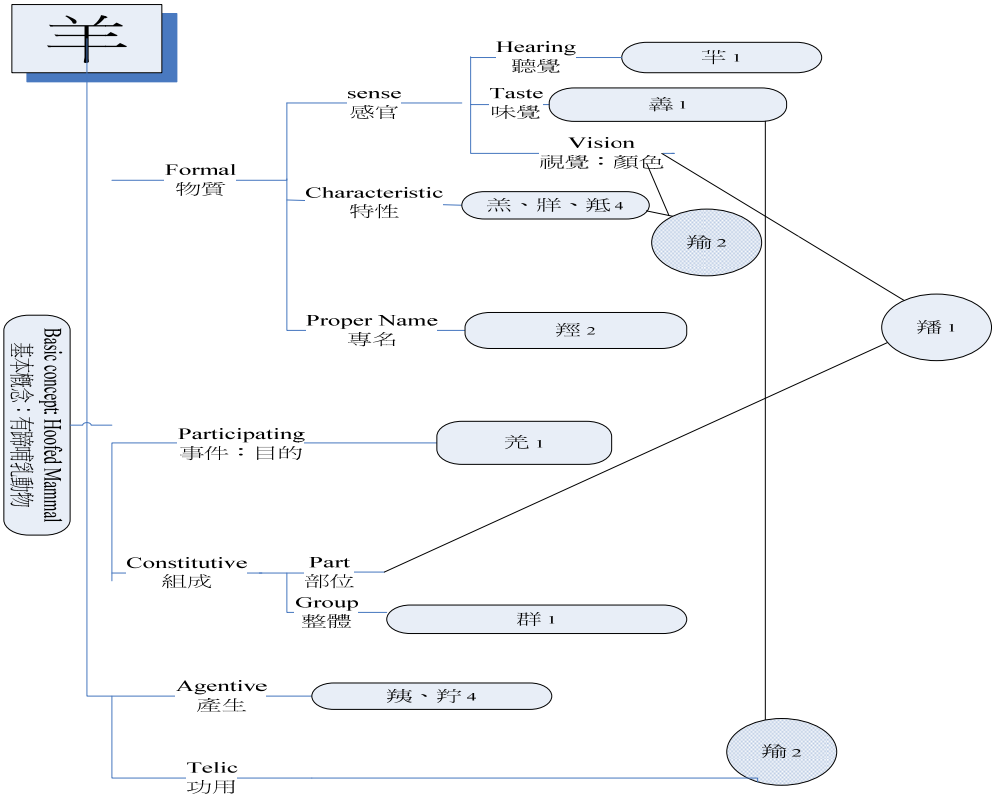

Figure3: The BOVID Domain Ontology Conventionalized by radical “羊”

(i)The cocnept cluster belong to “formal” mainly describe the color, sex, and age of bovid. Foe example, “{羊兆}，羊未足歲也。” **(bovid which is less than one year old)** expresses a concept involving the age of the bovid. “羶，黃腹羊也。” (**bovid which has yellow belly**) involves both the constitutive part of a bovid’s belly and its visual attribute. As in many other animal concepts, sex is also an important concept. For example, “**羭，牡羊也。”(male bovid) “牂，牝羊也。” (female bovid)**

(ii) A smaller cluster of concepts denotes events which bovid can be involved in and is classified as Participating. For example, “羌，西戎羊種人也。” **(shepherd in Xiyu)** The concept of a particular type of human is defined by referring to their relation with bovid, We categorize this word into participant-goal. Besides, ‘羌”combines two semantic symbols “羊” and “人.” Note that it could be argued that the basic concept should be HUMAN. However, as mentioned earlier, we made the commitment of describing the concept classification of SuoWenJiezi in this first study and will make adjustment after the complete ontologies are completed.

(iii)The concept cluster related to “birth” are classified as Agentive. For example, **“羠，五月生羔也。” (lamb born in May)** In addition, there are words related to castration, such as **“羙，騬羊也。”(castrated bovid).** We classify it into “Agentive” since it denotes how this kind of bovid comes to being.

3.2. Deer Domain Ontology Conventionalized byRadical “鹿”

There are only 28 concepts in the clusters of radical “鹿.” Ict is not very productive compared with the other three animal radicals. It should be observed that deers were not domesticated and hence has a much less linked to direct human experience in archaic Chinese society. This is

reflected by the fact that concept cluster conventionalized by the radical “鹿” belong predominantly to the “Formal” calss with only one controversial case of “Participating.”

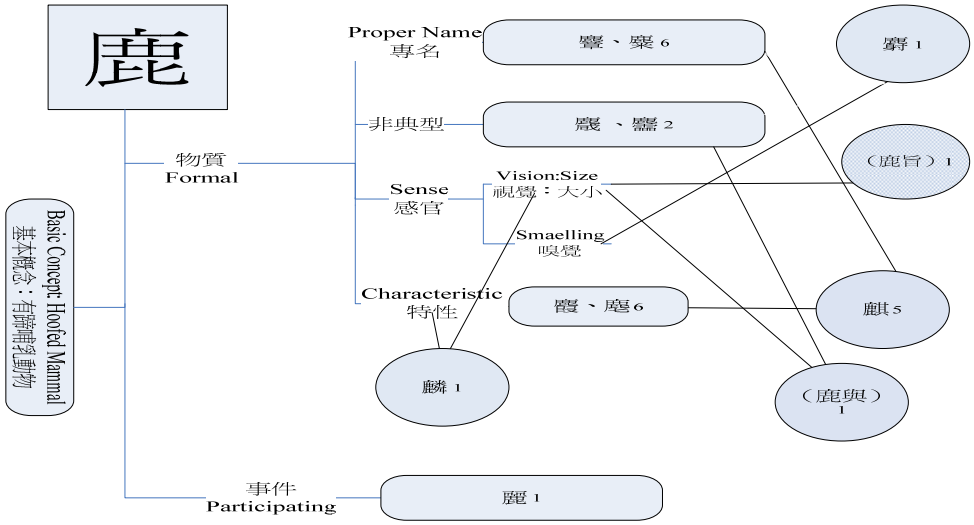

Figure4: Deer Domain Ontology Conventionalized by Radical “鹿”

(i)Most of the cocepts governed by the radical “鹿” are “Formal”, mostly in the “Proper name” and “Characteristic” classes. For example, “**麋，鹿屬.” “”**麋 is a “Proper name.” **“麀：牝鹿也。”(female deer)**. This describes the sex, in the “Characteristic” class. There are also concepts cross “Proper name” and “Characteristic” two categories. For example, “**麎，麋牡者.” “**麋” (moose) is a “Proper name,” and”牡”describes sex. This is a cross categories example.
(ii)There are also atypical category in deriving concept of radical “鹿.” For example, “**麙，山羊而大者.”(big goat)** Goat is not a kind of deer. So it could be a mis-classification, either by convention or by Xyu Shen.
(iii)There is a single example of concept derived from participating:“**麗，旅行也”(traveling)** The Chinese Dictionary explains traveling here as “travel with companion.”. The etymology and conceptual conventionalization cannot be clearly defined.

3.3. Cattle and Horse Domain Ontologies Conventionalized by Radicals “牛” and “馬”

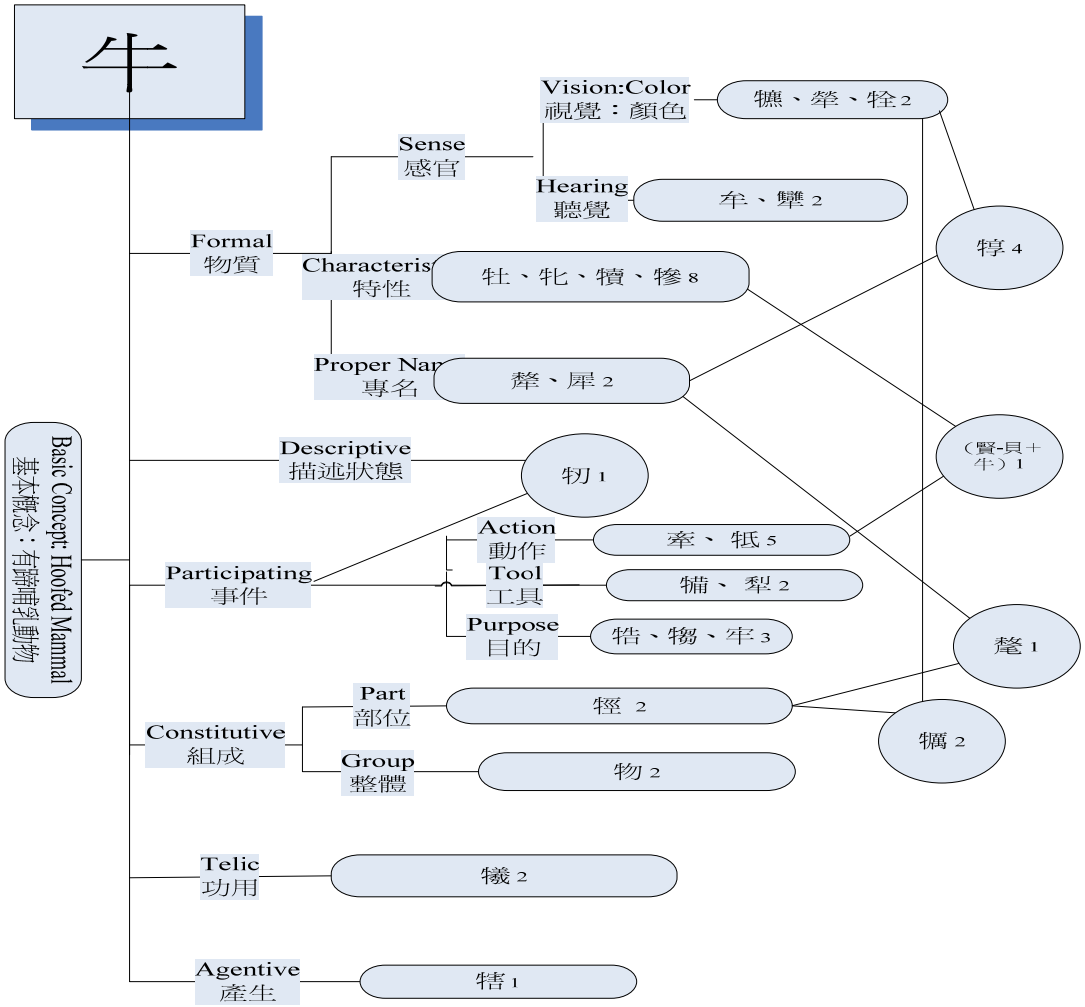

Figure5: Cattle Domain Ontology Conventionalized by of radical “牛”

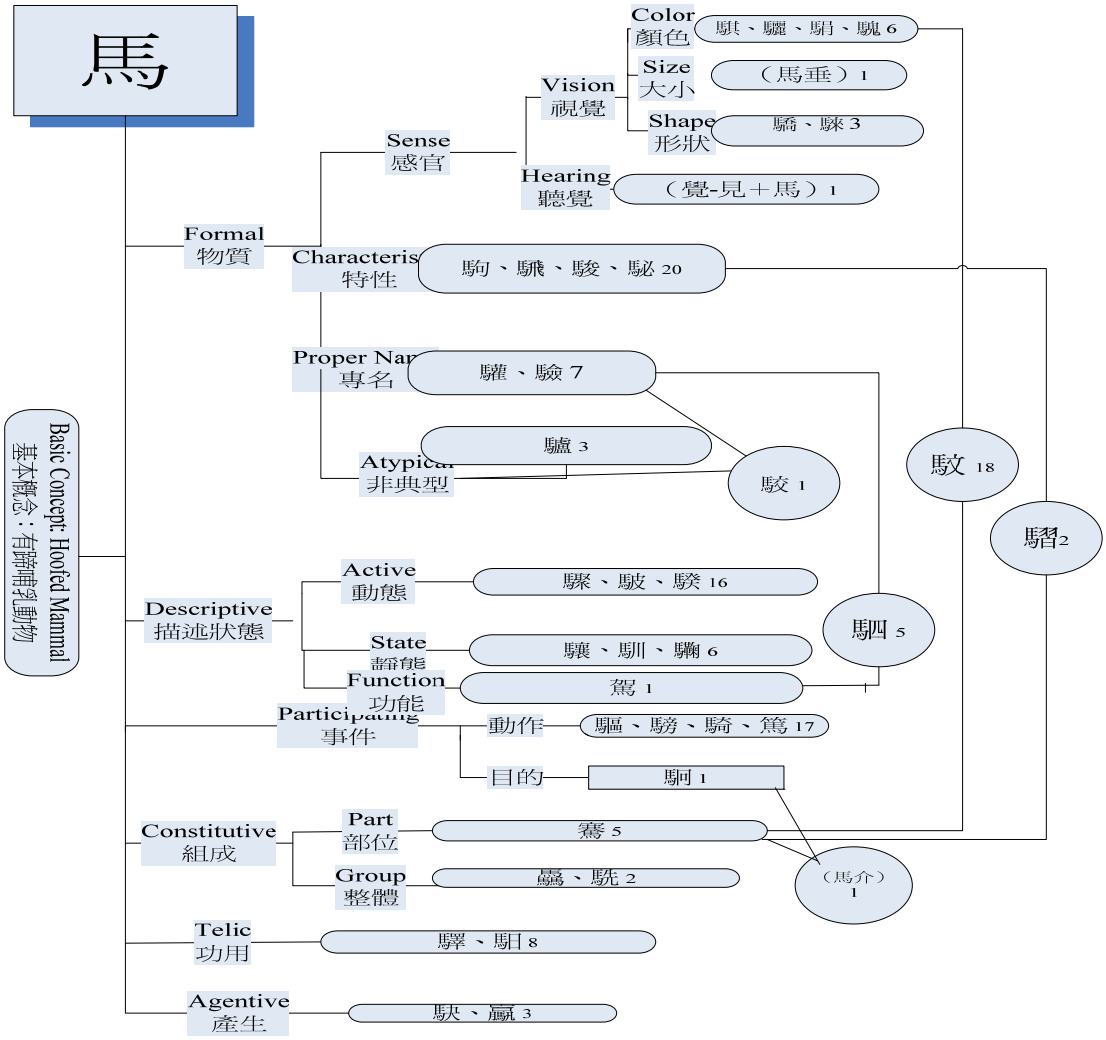

Figure6: Horse Domain Ontology Conventionalized by Radical “馬”

The Cattle and Horse Domain ontologies are much richer than the BOVID and DEER domain onotologies, which shows that these two animals are central to archaic Chinese society. The typical concepts are related to cattle being used to till the land and horses being used in transportation. Because of the close and rich first hand information, there are also descriptive events referring to experiences involves these two animals. For instance, the concept of ‘to scare’ and ‘being scared’ is actually represented by 驚 and derived of the drastic and vivid event of a startled horse.

4. Conclusion

In conclusion, we found that the definitions of the semantic primes of these four mammals with hoof are descriptions of the appearance of these animals:**(bovid represents the shape of (an animal with) four legs and a tail) (cattle represents the shape of (an animal with) horns which triangulates with a tail) (horse represents the shape of (an animal with) a maned head, tail, and four legs) (deer represents the shape of (an animal with) horns and four legs).** It is natural that the meaning clusters of concepts they derive belong to the “Formal” aspect, especially the “Vision” and “Characteristic” categories. Note that there are also many examples cross two categories, “Formal” and “Constitutive-Part,” reflect the basic concept definitions involved constitute parts.

Among the four semantic primitives, radical “deer” derives the least number of concepts. This reflects the knowledge of the speakers of the language as bovid, cattle, and horse, are known to be already domesticated whiles deers are not.

Among the three domesticated animals, cattle and horse are more similar to each other in how their drived concepts cluster and distribute. The distribution of these two radicals are both include “formal,” “Descriptive,” “Participating,” “Constitutive,” “Telic,” and “Agentive.” Bovid is different, perhaps owning to the fact that bovid was domesticated mostly for food while cattle and horse serve the functions of farming and transportation.

It is also important to note that religion and rituals do play central role in human experience at that time. These can be observed from the definitions of the derived concepts. We observe these functions from “telic” category and find that the occasion people use cattkle is when making a religious offering. For example, **“牲, 牛完全也.”(the whole cattle used for offering).** However, horses do not seem to be offered as sacrifice but can testify to a developed system of transportation, **“驛, 置騎也.” (Yi, a place wherehorses are posted).**

In conclusoin, our study of four hoofed anminals support our original thesis that the basic concepts as conventionalized by radicals represents a domain ontology of a cluster of concepts derived and marked by that radical. It is important to note that this conventionalized ontology reflects the human experience and knowledge at the time of conventionalizaiton. Hence we see that the domain ontology of domesticaed and non-domesticated animals differ from each other, while further distinctions can be made between food animial and labor animal. These direct experience and knowledge is reflected in the qualia used to derive these concepts.

5. References

Chou, Ya-Min (2005). Hantology: A Chinese Character-based Knowledge Framework and its Applications. Ph.D thesis, National Taiwan University.

Chou, Ya-Min and Chu-Ren Huang. (2007). Hantology: An Ontology based on Conventionalized Conceptualization. Presented at the Fourth OntoLex Workshop. Jeju, Korea.

Hsieh, Shu-Kai. (2006). Hanzi, Concept and Computation: A preliminary survey of Chinese Characters as a Knowledge Resource in NLP. Ph.D thesis, Universität Tübingen.

Pustejovsky,J. The Generative Lexicon, The MIT Press, 1995.

Xu, Shen. 121. 說文解字 ShuoWenJieZi (in Chinese).

Xu, Zhong-Shu, 漢語大字典 The Chinese Dictionary, Taipei, Jian-Hong Press, 1992

Wong, Shun-Ha and Karel Pala. 2002. Chinese Characters and Top Ontology in EuroWordnet. In Singh, U.N. (ed).Proceedings of the First Global WordNet Conference. Mysore, India.